Assignment 7

Write a python program to implement a Logistic Regression Classifier, using the following instructions:

    1. Load the PIMA Indian Diabetes dataset.

    2. Do a train test split of 75% and 25% of the data.

    3. Scale the data using Standard Scaler.

    4. Fit the Logistic Regression model on the training data.

    5. Do a prediction using the model on the test data and print the confusion matrix.

In [4]:
```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import confusion_matrix

# Load the dataset
file_path = 'diabetes.csv'  # Replace with the path to your dataset
diabetes_data = pd.read_csv(file_path)

# Split the data into features (X) and target (y)
X = diabetes_data.drop('Outcome', axis=1)
y = diabetes_data['Outcome']

# Split the dataset into training and testing sets (75% training, 25% testing)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=16)

# Scale the data using StandardScaler
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

# Fit the Logistic Regression model on the training data
model = LogisticRegression(random_state=16)
model.fit(X_train_scaled, y_train)

# Make predictions on the test data
y_pred = model.predict(X_test_scaled)

# Print the confusion matrix
conf_matrix = confusion_matrix(y_test, y_pred)
print("Confusion Matrix:")
print(conf_matrix)
```

```
Confusion Matrix:
[[116    9]
 [ 26   41]]
```

In [ ]: